DOCUMENT RESUME

ED 344 915 TM 018 245

AUTHOR Mengel, Bill E.; Schorr, Larry L.
TITLE Developing Item Bank Based Achievement Tests and Curriculum-Based Measures: Lessons Learned Enroute.
PUB DATE 92
NOTE 13p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; *Achievement Tests; Compensatory Education; *Educational Assessment; Elementary Secondary Education; *Item Banks; Multiple Choice Tests; National Norms; Norm Referenced Tests; Portfolios (Background Materials); Public Schools; *School Districts; State Programs; *Test Construction; Testing Programs; Writing Evaluation
IDENTIFIERS *Colorado Springs Public Schools CO; *Curriculum Based Assessment; Education Consolidation Improvement Act Chapter 1; Mandated Tests; Performance Based Evaluation; Rasch Model

ABSTRACT

The experience of the Colorado Springs (Colorado) Public Schools (CSPS) in using different assessment forms for different purposes is detailed. A small niche is reserved for national norm-referenced testing to fill the state achievement testing mandate and for Chapter 1 accountability. A second, and larger, part of the assessment program will be in the performance area. At the present, the CSPS district is beginning to develop performance assessments, starting with portfolio and writing assessment. A third dimension, the major focus of this paper, is curriculum-aligned levels testing. One approach will use a traditional format of assessment of basic (but not necessarily lower-order) skills with multiple-choice items, drawing content from the Rasch-calibrated item banks of the Northwest Evaluation Association. Construction of these District Achievement Levels Tests (DALT) is discussed. A second approach is through curriculum-based measurement (CBM) currently used in assessing oral reading, written expression, spelling, and computation. Problems that the district has addressed in DALT and CBM are described. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED344915

Developing Item Bank Based Achievement Tests And Curriculum-Based Measures: Lessons Learned Enroute

Bill E. Mengel and Larry L. Schorr
Colorado Springs Public Schools

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
BILL E. MENGEL
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

In a climate where everyone from parents, to school boards, to classroom teachers, to measurement experts, to the President, have the hots for 'alternative,' 'authentic,' or 'performance assessment,' it is somewhat risky to prepare a paper that supports the value of multiple-choice, objective or norm-referenced testing. Referring to this situation, Dennie Wolf has recently characterized it as "the romance of alternative assessment." As we have examined the alternatives we have chosen a middle ground, using different assessment forms for different purposes, and helping to conserve the scarce resources of the public schools. Instead of a "romance" with 'alternative' assessment we have pursued a simple, more natural, friendship.

In our district's assessment program a small niche is reserved for national norm referenced testing. In our system it has filled a state achievement reporting requirement of one grade level per school and serves as the required measure for Chapter 1 accountability. A second, and hopefully much larger part of our assessment program will be in the performance area. We have used future tense here because we are at the very beginning in the development of a performance assessments. At this point we are beginning to work with portfolio assessment and providing training and consultation in six-trait analytic writing assessment. We see these and other performance measures as most useful at the student and classroom levels and have yet to develop a plan as to how these might be used in a large scale district assessment program. Objective referenced, curriculum aligned levels testing is the third dimension of our assessment program and is the major focus of this paper.

Popular usage can distort words from their original meanings. In the assessment world, 'alternative' is such a distortion since, in a multi-dimensional assessment program, each form of assessment represents an alternative. The validity of each form is affirmed if it is congruous with the function(s) it is asked to perform. In most situations performance based measures are probably more appropriate to use in assessing students' depth of understanding. When greater breadth is required of an assessment, or when we are looking for greater objectivity, or easier aggregation of data is required, or economy of administration in time or money is a priority, then an alternative using a multiple-choice or other format using more narrow parameters may be desired.

This paper does not argue the relative merits of this or that form of assessment. That has been done to death elsewhere. We want to share one district's initial experience with two forms of assessment that are new to our experience. Both have been and are in use elsewhere. The first, and the

TM018245

2 BEST COPY AVAILABLE

one we will discuss in greatest detail, uses a traditional format of assessment of basic (not necessarily lower order) skills with multiple-choice items and drawing content from Northwest Evaluation Association's Rasch-calibrated item banks for construction. This approach has been used throughout the Northwest for a number of years. We think it is worth revisiting for those of you who have seen it before, and is worth visiting for the first time by those who may be looking for a way to match curriculum and instructional objectives to economic assessment.

Our second goal is to discuss 'dynamic' (our use of the word) assessment as described by Deno, Shinn, and others in their presentation of curriculum-based measurement (CBM). Our experience with CBM began several years ago when, as practicing school psychologists, we became increasingly dissatisfied with assessments that had little utility for the classroom teacher and which did not serve the end of more effective instruction. Neither of these discussions will be highly technical. Instead, they will try to give you the flavor of what we have done, the problems we have encountered, and in some cases overcome, and the lessons we have learned which may be useful to others as they strive to maintain a balance in alternative assessment.

Our multidimensional assessment program began with two developments, the reorganization of administrative functions in central administration with the arrival of a new department director, and a new state finance act. The latter, in effect, rewrote state accountability requirements relating to proficiency assessment and diploma certification. Our new director pushed us to examine prevailing assumptions about the adequacy of the then current unidimensional, national norm referenced assessment program (NRT) to serve multiple purposes.

For almost a quarter of a century the district had relied almost exclusively on the Iowa Tests of Basic Skills to perform multiple functions (with a brief and very unpopular flirtation with constructing our own criterion-referenced tests). These ranged from diagnosis of individual student learning difficulties, to instructional grouping and course selection, to school and district accountability. And all of this without critically examining the purpose for which such tests were constructed or the degree of match with students' cognitive levels or the curriculum. What we did know was that some of our students and schools got an uneven start and, according to test results, never caught up. But what can you expect in a world where abilities are distributed 'normally'? The district generally performed well on these tests so the assumption was that we were doing a decent job of educating our children.

Our current knowledge is a balance of experience, and what we read in the professional literature: NRTs can serve only a few narrow purposes and we can no longer try to stretch them to meet all the 'new' desirable ends of assessment. As we have begun to move toward a multidimensional assessment system we have encountered many challenges. It is the primary purpose of this paper to share those challenges and our ongoing attempt to find solutions to them in the hope that the 'lessons' we learned and continue

to learn may be of help or interest to others engaged in the process of changing assessment programs.

Description of DALT and CBM

The District Achievement Levels Tests (DALT) are our standardized tests. They are objective referenced, multiple choice tests that provide information about students' progression toward proficiency in those skills that represent the district's basic skill objectives. The items are matched to curriculum objectives, are of known difficulty based on Rasch calibration, and are placed in overlapping levels of difficulty. Students are given the level of test that is most representative of their current levels of skill development. Each level has a large enough number of items, matched to district objectives and within students' current range of academic development, to give us valid and reliable information about mastery of the district's curriculum objectives. In turn, this provides more useful instructional direction because of increased confidence in the measure.

We use CBM measures to assess students' fluency in oral reading, written expression, spelling and mathematical computation. It is most appropriately used to monitor student progress in developing the basic fluency necessary to higher order skills in reading, language arts, and mathematics. Because it is actually a performance measure it overcomes some of the contrivedness for which objective referenced tests are often criticized. It uses actual curriculum source material for the measures. CBM measures can be administered frequently, twice a week or more, because of an almost unlimited number of alternate forms and brevity of administration, with modules of no more than three minutes in length. This allows for dynamic decision making about instructional effectiveness that is one of the properties that is so attractive in other performance measures. It can be used with or without district norms. We have found CBM to be a useful element in special education and Chapter 1 decision making, regular classroom monitoring of student progress, and as useful measures in our school improvement plans.

Problems Faced-Lessons Learned

Our problems fall into definite but interrelated categories. We have identified these six general classifications as: inertia, political, resource, technical, administrative, and utilization. These categories seemed to hold whether we were considering the DALT or CBM. It is necessary, however, to talk about them separately as they manifested themselves in different ways. For the sake of organization we will treat them independently, discussing first problems and solutions in instituting the DALT. A briefer, less categorical, discussion focusing on CBM will be addressed last.

District Achievement Levels Tests

Inertia

We first challenged inertia. The American Heritage Dictionary defines 'inertia' as "the tendency of a body to resist acceleration;" or "the tendency of a body at rest to remain at rest or of a body in motion to stay in motion in a straight line unless disturbed by an external force." All three facets of this definition seemed to be represented by the entrenchment of an NRT system which spanned a quarter of a century and with which there was general acceptance or satisfaction. It was seen as slightly intrusive but not very significant in disturbing the 'routine'. We had educated the consumers of NRT data to believe that it provided some useful information about the progress of the district and even of individual students. Familiar, if sometimes misused, metrics were used, further lulling administrators, teachers, and parents into a misplaced comfort level with the national NRTs. Since tests were viewed by many of them as a necessary evil, it was pushing them a bit far to suggest that data could be helpful for instructional decision making, and a bit further still to suggest that they needed a new kind of test, with unfamiliar measurement terminology, to make appropriate decisions at the classroom or student level.

Initial progress in resolving this problem began with informal dialogue between the director of our department (Planning, Evaluation, and Measurement (DPEM) and the head of the Department of Instruction (DI). These conversations about the role of assessment in driving and/or responding to the needs of instruction paved the way to avoid the potential political pitfalls which might have lay ahead of us. Also instrumental in overcoming inertia was the passage of a state accountability act which rendered the NRT inappropriate for the purposes it demanded

Political

District inertia had the potential for spawning political problems. When one political entity sees the need for change and communicates that to another political entity. These entities were represented by DPEM and DI. DPEM's position was that we could not meet new accountability mandates that called for the establishment of proficiencies and continue to use a measurement system that was related to the curriculum only tangentially. Fortunately, the leadership in DI had become concerned about what they described as "achievement gaps" among and within the fifty-three district schools and the problems of integrating evaluation with curriculum and instruction. The adoption of our new assessment program, with its array of performance, objective referenced, and national norm referenced options, coincided with DI's awareness of the need to have a purpose driven assessment system. They were eager to participate in discussions of what that assessment program would look like and together we established the Curriculum, Instruction, and Assessment (CIA) advisory group to review common issues.

One of the first decisions was to move quickly to overhaul the current assessment program with its almost total dependency on NRTs. Since our director had come from Portland Public Schools he was intimately familiar with their levels testing program. After discussion with DI it was decided that the best and quickest place to get a 'payoff' on our initial efforts was to build a curriculum aligned objective referenced assessment program that would replace the NRTs and could be tied to designated proficiencies. Northwest Evaluation Association (NWEA) was invited to present its item bank approach to building curriculum aligned objective referenced tests. DI selected groups of teachers to align questions from those banks with district objectives and to distribute them across the various levels.

We were also faced with selling the program to the community. We have generally addressed this problem through school building accountability groups. Again, the major problem is the entrenchment and perceived legitimacy of the NRTs and the 'suspect' nature of building a test that measures the district's objectives rather than using more 'respectable' national measures. Our solution has been to discuss appropriate use of assessment related to purpose and to examine the technical abilities of the DALT and the NRTs to fulfill those purposes. While we have not been able to fully accomplish it, our aim is to meet with accountability groups in every school building to obtain their support.

We expect more problems from state graduation requirements mandating evidence of proficiency in designated areas. Our district has been slow to conform to this aspect of Colorado's accountability rules. We anticipate acute political 'interest' when this finally takes place. The DALT and other IRT based proficiency measures, already in place, will provide part of the solution to this political problem.

Resources

Dwindling resources is a serious problem, especially for those programs which are perceived to reside outside the pale of the bare essentials. We were helped in our initial quest for resources by the passage of the 1988 state finance act which tied continued support to evidence of progress toward state mandated achievement goals. This meant changing the way we assessed student progress so that those measures would be related to instructional input. Our local Board of Education readily supported our request for start up funds and we were able to build and publish our first edition as well as purchase the necessary scanning, scoring, and reporting hardware and software.

We are looking for alternative funding and cost-cutting measures. This spring, as the time approached for reviewing the DALT for revision we were again resource short. We found that changing the format from a multilevel edition to single level tests we could save two-thirds the cost. The new Macintosh software just released by NWEA helps by formatting the test in far less space than was required when we had to reformat from the PC version and, as a spinoff advantage, the print quality of the items is much sharper than in the PC version.

Test revision increases costs. As long as their is a commitment to frequent revision of the tests there will be some significant costs. The advantage is that the measures can be kept very current, reflecting curricular change and aligning the test ever more closely to instructional goals. Frequent revision also allows for the removal of technically inadequate items, improving the validity of the measure as a whole.

We hope to divert significant resources from the NRT program. Anchor studies with NRTs may make variances possible and free up some funds from categorical sources which would otherwise be devoted to the purchase of commercial tests. There may also be possibilities of generating grant money from innovative assessment proposals.

Technical

We have used "hired guns" for technical problems. Technical difficulties were largely overcome through the consultative help received from NWEA and Microsystems for Education and Business (MEB) and software developed by MEB. The challenges consisted of building the levels tests around school district objectives, recalibrating the difficulty of math items when used with calculators, developing locator tests to place students in appropriate levels, deciding on an appropriate print format, obtaining a software package that modified the scanning, scoring, and reporting to our specifications, reformatting items from the PC based item bank to the Macintosh, anchoring the DALT results to a NRT, and realigning items to our curriculum for the second edition of the DALT.

NWEA helped us focus on our curriculum objectives. Teachers and administrators from the DI were brought together in a retreat setting for a two day workshop. Together they looked at curriculum guides and basically settled on four major goals in reading and eight in mathematics. They then divided into small groups to select items matched to specific objectives within given difficulty ranges. Each group had access to the item bank and was able to build levels tests on a PC through the NWEA test building software.

Calculators presented a problem and an opportunity. The math construction group decided that they wanted students to be able to use calculators in the solution of math problems at the secondary levels. The NWEA items had not been calibrated with the use of calculators. Therefore, it was necessary to choose similar items to those selected by the math group, administer them to an experimental and a control group, and recalibrate the test items based on the differential performances of the two groups.

Calculators can affect item difficulty dramatically. As expected, operations involving whole numbers became easier, by an average of 10 or 11 Rasch Units (RITs). Operations using decimals when presented in horizontal format became easier, ranging from 12 to 18 RITs lower in value. Estimation items became slightly easier, by two or three RITs, and probably ceased to be estimations. Items related to the numeration and geometry

objectives generally became more difficult for the calculator group by five or six RITs. Other objectives showed little calculator influence with a mild but probably not significant effect of making the items more difficult.

NWEA and MEB developed locator tests and modified these at our request to reflect our district's instructional vocabulary. The locators continue as the most unpopular feature of the DALT because of their relative difficulty but are necessary for students' initial levels placement. Subsequent use of the locators is not necessary as the answer sheet printing program estimates the students' levels from previous DALT performance. In the pilot we hand scored all the locator tests ourselves, learning another valuable lesson about overextension. In the fall we provided a key to teachers and allowed them several options, including student self-check.

We had a number of technical problems with our print and layout formats. The first was that the item banks were, at the time, available only on the PC, and needed to be reformatted on the Macintosh to provide the quality of print that we wanted. This meant that all figures, charts, or illustrations had to be redrawn on the Macintosh. We also tried to conserve space in our item layouts which caused us to vary answer choice and item presentation in an inconsistent manner. Our item analysis indicated that this interfered with the validity of some items and led us to using a consistent format in our newest revision. The new Macintosh version of the NWEA item banks largely overcomes the difficulties just noted, formats the test with an eye to economy of presentation, and allows editing capabilities to suit local needs.

We also encountered problems with the test booklet production. The creation of a multilevel edition seemed to be the most appropriate way, from an administrative point of view, to produce the tests. We would not have to worry whether we had sent enough of each level to each school. Our worst nightmare occurred when the books began falling apart. Other problems occurred when we discovered that some books had been collated incorrectly. During the summer we worked with our print shop in rebinding the tests with a plastic spiral and trying to ensure proper collating of the pages. While the results were much better, we still found that some students have a large capacity for destruction and that many of the spiral bound books "came apart" after minimal use. A cost analysis for the 1992 revision showed that we could save at least two-thirds the cost of printing by producing single level editions. These will be printed on 11 by 17 inch paper and stapled in the center just as are single level NRTs which have proven to be durable. Also, if mistakes are made in a given level they can be addressed in a more cost effective manner. Examination of levels used last year have guided us in the production of appropriate numbers at each level. The single level also places the responsibility in the teachers' hands for the student taking the correct level and eliminates the possibility that the student will select the wrong level.

The NWEA software package provided formats for alpha and rank reporting at the district, school, and classroom level. Scores were reported as RIT and POP (a standard score with a mean of 50 and standard deviation of 10) scores and as curriculum goal mastery for each student. A parent report is

generated which reports these same scores. It also includes a short explanation of each of our curriculum goals being assessed. We had it modified to include a local percentile as a way of transitioning from our traditional score reporting formats. Arrangements were made with MEB to interface with our student information system so that DALT scores could become a part of the student's test history file and to enable us to disaggregate scores by ethnic and gender for state reporting and internal use.

We commissioned MEB to do an equivalence study so that we would be able to relate performance on the DALT to the NRT. This helps to allay fears that we will somehow become complacent and ingrown with our performance on the district tests. It is also useful when conveying information about students who leave our district. A third planned use is to submit the results to the state for waivers from requirements for NRT reporting and to allow the DALT to be used more widely for reporting in categorical programs. This permission has already been won for ELPA category identification.

Realignment of items to more appropriate levels was necessary after examining the item analyses from the fall administration. We found that, in general, reading items retained the same RIT values in actual administration in our district as they had in the item banks. Following the lead of states like Illinois and Michigan, teachers will be encouraged to include some longer reading passages in the new revision.

Math items tended to fluctuate more widely. We saw this as largely a product of two factors. The first was a math curriculum that was divergent from districts represented in the item bank in the sequence of presentation of math objectives of varying cognitive levels. We hypothesize that the second reason was the use of the calculator at secondary levels, whose effect had not been totally compensated for in the initial corrections. Releveling the tests based on our district's fall performance, and eliminating the few items which did not perform well from a technical standpoint, will help to tighten the alignment with our own curriculum.

Administrative

Problems with details of administration are endemic to any form of assessment. We had many decisions to make. The first was to recruit a number of schools to pilot the spring, 1991, administration. We sent informational materials to all schools inviting their participation. After meeting with a number of individual school staffs twelve elementary and all nine of our junior high schools agreed to participate. We also gained their cooperation in the concurrent administration of ITBS in grades 3 through 9. This allowed us to collect the data we would need for the anchor studies at each grade level. Explanations of our reasons for changing the program, creation of locator tests for each level along with directions for giving, scoring, and interpreting them for appropriate levels test placement, creation and distribution of multilevel test booklets, preprinting answer sheets by classes, scanning the answer sheets in a timely manner, and reporting the data in a way that could be understood by parents, teachers,

and building and central administrators, were among the many problems we faced and overcame in the pilot.

In the fall administration we faced most of the same problems but on a larger scale. We tried the approach of dissecting the tasks involved in the administration and reporting of the DALT. With each phase we developed a separate written communication to teachers, administrators, and parents. The first communication covered the reasons for change in the assessment program, the advantages of the level system, and an anticipation of what was yet to come. The second dealt with the locator tests and how to deal with them. The third explained the DALT administration including how to identify the level on the answer sheet, sharing materials among teachers, appropriate administration practices and return of used answer sheets to us for scoring. A fourth communication dealt with the return of various reports and their interpretation. This parceled approach seemed to work well with those who bothered to read them. For those individuals who usually fly by the seats of their pants there were the usual problems. These were usually passed on to us in the form of unidentified answer sheets, levels indicated on answer sheets which did not appear to be the level the student had taken, wrong sections of the answer sheets used, and other opportunities for creative problem solving on our parts. Despite these problems we found that we were generally giving schools a two or three day turnaround on their results.

Innovations are often hard to explain, particularly when the vocabulary is strange or new. It is tough to get the level of understanding of concepts, such as 'RIT scores' and 'levels testing,' in the professional education and lay communities that is needed to ensure widespread support. During the time that we planned our pilot, and even after its full implementation this fall, it has been hard to get audiences except in those places that have a natural inclination to be excited about, or hostile toward, change. In those cases we have been invited to address faculty and parent groups. Our experience is that these opportunities greatly facilitate the understanding and acceptance of these measures. This is particularly true when we are able to frame them in the context of a multidimensional assessment program. Our recommendation is that these opportunities need to be a deliberate and planned part of any innovation involving alternative assessment.

Utilization

The final problem has been one of utilization of information. Here we are confronted with the same dilemma that we face with any form of traditional assessment, getting the consumer, usually the teacher, to use the information to enhance instruction. As proponents of "authentic" assessment rightly claim, teachers are more apt to use data that occurs in a naturalistic setting. The DALT overcomes some of the ambivalence by linking itself more tightly to the curriculum than a traditional norm referenced test. As we are able to spread that message thoughtful teachers are more likely to use the information in modifying instruction. In the meantime, we have found that Colorado's new and more stringent requirements for measurable accountability in our school improvement plans and the requirements for

the establishment of proficiencies at given grade levels blend nicely with the continuous growth scale provided by the DALT RIT scores. Requests by the overwhelming majority of our elementary and junior high schools to do spring assessment with the DALT is testimony of its acceptance for at least those limited purposes.

As we have greater opportunities for meetings with instructional staff concerning utilization of DALT data we believe they will see it as a useful adjunct to other less objective and quantifiable measures. The assurance that there are an adequate number of items within restricted cognitive ranges makes the DALT far more useful for analyzing the performance of individuals and groups of students. As we do more detailed analyses around ethnic and gender categories we may also be able to find instructional clues that can be useful in reducing the gaps that we see.

Curriculum-Based Measurement

Our CBM experience began in 1988 while we were still assigned part time as school psychologists. We began using it in our practice as an effective means to identify and diagnose problems in reading, written expression, spelling, and math computation, and to monitor the effectiveness of instructional interventions. Almost immediately, we were faced with a whole series of problems with implementation that fit in the same categories as those used in our discussion of the DALT. Rather than treat them discretely, as we did with the DALT discussion, these will be presented as a collage which may give you a better feel for the process and problems which occurred in its implementation.

Resistance to change came principally from two groups, our fellow school psychologists and special education practitioners who were reluctant to give up established measures which they used in developing their Individual Educational Plans (IEPs). Many psychologists and educators were and are captivated by the concept of aptitude-treatment interactions (ATI) which fits the idea of a diagnostic-prescriptive model. Unfortunately, the success of any treatment can be judged only after the fact. Most assessment techniques occur too remotely from the intervention to supply the educational practitioner with useful information. This does not, however, disturb the comfort level of those who have "always done it that way" or who wish to maintain a professional mystique. Wechsler tests and Woodcock-Johnsons continue to fit that model and are used to fit the aptitude-achievement discrepancy required by many states for identification of handicapped youngsters.

To overcome the inertia and what turned out to be political issues, seen as threats to traditional psychologist roles, we attempted to bombard our targets with research studies supporting CBM as a way of monitoring the treatment. Additionally, we visited and sent others to Pine County and the Minneapolis Public Schools where CBM had been used in special education decision making for many years. We sent principals, special and regular

education teachers, and psychologists to summer CBM institutes at the University of Oregon, and we brought Dr. Mark Shinn from the University of Oregon to our district for meetings with teachers, principals, special education administrators, and psychologists. We even recruited two outstanding doctoral students from the Oregon program to work as psychologists in our district.

As a result of our efforts, several schools requested pilot programs where every student was tested three times during the year and student progress was monitored using CBM. Simultaneously, we decided to build school district norms by creating random samples of 200 students at each grade level from 1 though 10 and testing each of them in the winter, spring, and fall of 1990. This made it possible for us to apply to the state for a waiver of the usual requirement in demonstrating an aptitude-achievement discrepancy for handicapped eligibility. We requested and received permission to substitute an achievement-achievement discrepancy, with the locally developed norms as the reference point.

Through the support of our local special education unit and the state special education department we were able to receive financial support for the development of testing materials and the materials necessary for progress monitoring. We believe that progress monitoring with CBM is a form of dynamic performance assessment which allows the teacher to gauge the effect of her instruction on a limited goal, that is, fluency in one or more of the basic skill areas as measured by real reading within curricular materials, real writing tasks, etc. While performance assessment purists might judge the vehicles as contrived rather than 'natural' these real performances enable the teacher to make frequent changes during the course of her instruction, and then, on an ongoing basis, judge the effectiveness of those changes.

Acceptance of CBM has not been universal but is gaining in popularity, especially among special education and Chapter 1 teachers who are interested in measures of instructional effectiveness. We continue to promote it through the offering of in-service education and have received state grants to fund these activities. It is being used in many of our schools, particularly in primary grades, as a measure of objectives in school improvement plans. Data may be summarized by grade to indicate the amount of average growth taking place between pre and post measures. Some schools have chosen to administer measures on a monthly basis in order to provide periodic snapshots of class or school growth in fluency. While it can be used in special education certification there are few of our psychologists using it for that purpose.

While there are those within the measurement community and even within our local education community who would criticize CBM as focusing on objectives that are too narrow or superficial, there is a strong research base indicating the importance of fluency as an important tool to promote the development of higher order abilities. Again, it has not been the primary purpose of this paper to promote or defend assessment paths taken but rather to share our experiences along those paths.

Conclusion

My old grandpa was a lay preacher. There is a story which has been retold in the family many times. It seems one Sunday morning he stood up to preach, paused, and said, "I'm not very much interested in this topic and I don't think you would be either." And he sat back down. That's about how we feel now after having waded through the writing of this paper. However, if you are still with us, we hope that you have found something of use, if not of interest, and that it may provide you with some ideas of alternative assessment in addition to the performance or 'authentic' approaches now receiving widespread attention and acclaim. The alternatives described here are not meant to displace any other forms of assessment but to supplement them. The driving consideration for us is always, for what purpose do we assess? The answer to that question should drive the form of assessment.